A DASHBOARD TO MONITOR HOUSING MARKET

BY

KEE LI YAP

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF COMPUTER SCIENCE (HONS)

Faculty of Information and Communication Technology
(Kampar Campus)

JANUARY 2020

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: A Dashboard to Monitor Housing Market

Academic Session: JANUARY 2020

I KEE LI YAP

(CAPITAL LETTER)

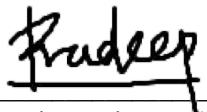
declare that I allow this Final Year Project Report to be kept in

Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

(Author's signature)



(Supervisor's signature)

Address:

6, Jalan Melor,
Taman Sri Sekinchan 2,
45400 Sekinchan Selangor.

Dr Pradeep a/l Isawasan
Supervisor's name

Date: 24 APRIL 2020

Date: 24 APRIL 2020

A DASHBOARD TO MONITOR HOUSING MARKET

BY

KEE LI YAP

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF COMPUTER SCIENCE (HONS)

Faculty of Information and Communication Technology
(Kampar Campus)

JANUARY 2020

DECLARATION OF ORIGINALITY

I declare that this report entitled "**A DASHBOARD TO MONITOR HOUSING MARKET**" is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : 

Name : KEE LI YAP

Date : 5 MARCH 2020

ABSTRACT

This project is to develop a dashboard to monitor housing market. By delivering dashboard to provide at a glance information to the user. Solving the problem of difficult analysis while switching on multiple web applications. At the same time provides data for similar projects that are going to conduct in the future. The focus of the initiate project would be in Kampar. As to collect non-city data, it would bring more chances to the trend of research and development in Kampar area. Therefore, collecting data is costly and time consumed, most of the organization would not be conducting the similar analysis without resources that are ready to be accessed, especially when the main services of the organization is not on delivering data analysis. Besides, delivering dashboard having an unknown return value to the organization, develop a dashboard without resources that are ready to be accessed would be harmful and risky to an organization. This project would be a “white rat” to the market, to understand whether there are demands of delivering this type of services. Kampar, as a small town in Perak, with limited data provided on the internet, is one of the obstacles throughout conducting this project. Lastly, this project would be involved in data mining and data analytics to conduct a meaningful insight to users of the dashboard. Targeted user of the dashboard would be renters, property investors and existing landlords.

TABLE OF CONTENTS

TITLE i

DECLARATION OF ORIGINALITY ii

ABSTRACT iii

TABLE OF CONTENTS iv

LIST OF TABLES vi

LIST OF FIGURES vii

LIST OF ABBREVIATIONS ix

CHAPTER 1: INTRODUCTION 1

1.1 Problem Statement 1

1.2 Motivation 1

1.3 Project Scope 2

1.4 Project Objectives 3

1.5 Impact, Significance and Contribution 3

1.6 Background Information 4

CHAPTER 2: LITERATURE REVIEW 6

2.1 Introduction 6

2.2 Data Source 6

2.2.1 Mudah.my 6

2.3 Existing Dashboard 7

2.3.1 MarketMinder 7

2.3.2 Lancashire County Council Housing Dashboard 9

2.4 Factors affecting rentals of residential apartments 10

2.5 Regression Models 11

2.5.1 Elastic Net – Lasso 12

2.5.2 Radial Basis Function 13

CHAPTER 3: SYSTEM DESIGN 15

3.1 Overview of the Dashboard 15

3.1.1 Use Case Diagram 15

3.1.2 Activity Diagram 16

3.2 Timeline 17

CHAPTER 4: PROPOSED METHOD/APPROACH 18

4.1 CRISP-DM Approach 18

4.1.1 Business Understanding 18
4.1.2 Data Understanding 19
4.1.3 Data Preparation 20
4.1.4 Modelling 20
4.1.5 Evaluation 21
4.1.6 Deployment 21
4.2 Tools 22
4.3 User Requirement 23
4.4 Hardware and Software Specification 24
CHAPTER 5: VERIFICATION PLAN 26
CHAPTER 6: CONCLUSION 28
6.1 Novelties, Contributions and Problems Encountered 28
6.1.1 Open Source Dataset 28
6.1.2 Factor that Affecting Kampar Residential Properties Renting Price 28
6.1.3 Prediction on Kampar Residential Renting Price 30
6.1.4 Purposed Dashboard Overview 32
6.2 Personal Insight and Future Work 36
BIBLIOGRAPHY 38
APPENDIX A REQUIRED DOCUMENTS A-1
A.1 WEEKLY REPORTS A-1
A.2 POSTER A-13
A.3 PLAGIARISM CHECK RESULT A-14
A.4 CHECKLIST FOR FYP2 THESIS SUBMISSION A-17

LIST OF TABLES

Table Number	Title	Page
Table 1-1	Question to be answered by the dashboard.	2
Table 2-1	Factors that affect rental prices in Pune, India.	11
Table 4-1	Hardware specification	24
Table 4-2	Software specification	24
Table 5-1	Visibility testing on dashboard page	26
Table 5-2	Usability testing on dashboard page	26

LIST OF FIGURES

Figure Number	Title	Page
Figure 2-1	Market grade dashboard provided by AirDNA	8
Figure 2-2	Airbnb house renting performance dashboard by requested region developed by AirDNA.	8
Figure 2-3	Lancashire county council median house prices dashboard for year ending September 2018.	10
Figure 2-4	Definition of Lasso	12
Figure 2-5	Definition of Elastic Net	13
Figure 2-6	Architecture of Radial Basis Function Network	13
Figure 3-1	Use case diagram for the proposed dashboard	15
Figure 3-2	Activity diagram for the proposed dashboard.	16
Figure 3-3	Timeline for FYP 2.	17
Figure 4-1	CRISP-DM	18
Figure 6-1	Uploaded dataset on Kaggle	28
Figure 6-2	Correlation coefficient between factors and house renting price	29
Figure 6-3	Correlation coefficient between factors and apartment renting price	29
Figure 6-4	Correlation coefficient between factors and room renting price	30
Figure 6-5	Elastic Net Lasso, prediction model on renting prices	31
Figure 6-6	Radial Basis Function, prediction model on renting prices	31
Figure 6-7	Kampar housing dashboard in FYP 2	32
Figure 6-8	Exact location of each data	33

Figure 6-9	Average house renting price	34
Figure 6-10	View regions	34
Figure 6-11	Select region	35
Figure 6-12	Furniture and facilities radar chart	36

LIST OF ABBREVIATIONS

IMD	Index of Multiple Deprivations
CRISP-DM	Cross-Industry Standard Process for Data Mining
API	Application Programming Interface
KPI	Key Performance Indicator
RBF	Radial Basis Function
IDE	Integrated Development Environment

CHAPTER 1: INTRODUCTION

1.1 Problem Statement

The difficulty of analysing information from site-to-site.

Inconsistent information can easily be found in web applications. For instance, most of the residential property providers such as web applications did not own its property data. The objective of most residential property web applications is more focus on providing a platform to users to share their information. Data will be uploaded by their users. Therefore, inconsistent information will be occurred in web applications due to different set of data may not be entered by the same person. Besides, the communication between residential property web applications did not limit the same users to spam information on different web applications. This would be time-consuming for users to identify and distinguish the redundant information.

The vague information delivered by web applications.

While investigating the existing web applications. It resulted in most of the agents did not provide the exact location of the building on a map. It is important to have visualisation rather than word descriptions. 'Human brain can process information using graph or chart to understand complex data easily compare to bunches of description presented in words' (Anon., n.d.). It is better to visualise the data listed in web applications by agents. For instance, instead of stating "5 minutes driving distance from UTAR", a visualised map would tell the users as much information as possible.

1.2 Motivation

To provide an "at a glance" data analysis, a dashboard is introduced to focus on conveying fascinating information to catch this fast pace era's attraction. Collecting data from the existing web applications, however, would make this project possible. The developer believes, if it is possible to make the most use of the existing resources to create something that would return value to the community, it would return unpredicted results. Since the project would be costly for an organization to take the risk to research and develop with the vague of knowing the return value, the project

would be contributed on asking whether there are demands of this type of services in Malaysia market.

Besides, developing a dashboard would need a batch of data. There are limited data can be delivered from web applications. This project would be involving in collecting data and providing ease of access to the complete dataset. This would help the future projects which are working on the same type of study to have more opportunities to create something meaningful and gain more return to the community.

1.3 Project Scope

The initial dashboard development will be limited on monitoring renting market in Kampar area. There are a lot of hypotheses arguing the relationship between rental prices and purchasing prices, however, purchasing prices would not be considered in conducting the researched. This is mainly because the target audience for this project may like to see the solution on how to increase the renting value, whether on new purchases or currently owned of the residential properties in Kampar. Therefore, the dashboard would be focus on monitoring rental market instead of purchasing market to understand the value from analysing rental prices.

Throughout the development process, the developer would be focus on the existing attributes conveyed by web applications. Despite this may reduce the time to perform data collecting but this would increase the complexity to combine information from various web applications and to perform data cleaning. Besides, due to the limited data provided from web applications, the developer would consider conducting survey and distribute to students from UTAR, Kampar campus. This solution of expanding data size would help to find out the third business question that is stated in Table 1-1. (Klipfolio.com, 2019) Dashboard is fuelled by business questions. A successful dashboard should be involved with business questions to help in solving real life problems. Basic business questions proposed in the project is expected to have the ability to answer the questions listed in Table 1-1.

No.	Question Description

1.	What is the residential property renting price range of different type of unit?
2.	What are the factors affecting Kampar residential property renting price?
3.	Which are the popular regions that are always being rent in Kampar?

Table 1-1 Question to be answered by the dashboard

1.4 Project Objectives

To investigate the core factor that affecting rental prices of residential properties in Kampar.

Information gathered from different application and visualised on a single interface would improve the experience of monitoring residential renting information in Kampar. The developed dashboard would help the property investors and existing landlords in Kampar to improve their renting services and deliver higher value of their residential properties as well as helping renters to rent at a reasonable price. Therefore, they would get the big picture of what is happening in Kampar housing market by monitoring the dashboard. (Klipfolio.com, 2019) A successful dashboard would convey the most important message to the users. From where user can gain insight at a glance.

To develop a dashboard to monitor Kampar housing market.

The dashboard development throughout this project will follow the definition of dashboard defined by Stephen Few. This is because dashboard can easily achieve at a glance of data analysis, in order to solve the problem of constantly switching web applications for comparison and analysis. It would provide a better experience of monitoring Kampar housing market.

1.5 Impact, Significance and Contribution

The final aim of this project is to provide a quick view on Kampar housing market for the renters, property investors and existing landlords. The dashboard would help to monitor Kampar housing market at a glance through providing graphic visualisation that might inspire the investors on developing better accommodation which are best suit the market demand or purchasing and improve the property value to gain high

return on investment. For example, investor might notice that residential properties that are nearer to university have higher renting price compare to others. On the other hand, renters and landlords can estimates the best price to rent the unit through information gained from the dashboard, it can be higher or lower than the average renting price. Besides, spending time on collecting data from multiple web applications to develop a dashboard is costly and having an unknown return values for an organization. Therefore, the result of the project can tell whether dashboard is the tool for the market and tell the industry whether the market need it as one of a tools to simplify information gathering, at the same time, helping the organization to reduce risk on developing same type of tool.

Developing the dashboard would need a complete dataset. Currently, data of Kampar residential properties, whether is complete distribution of residential properties or a complete dataset of the research cannot be easily found on the internet. The result of the project, therefore, can provide a completed dataset, to assist the upcoming research and development on Kampar housing market. This would attract more projects to involve to the similar aspect of the project, since it provided the ease of access to the data collected by this project from multiple web applications and by conducting survey.

1.6 Background Information

(Singla & Bendigari, 2019). Nowadays, information of residential properties can be easily accessed by any individuals from internet. Although information conveyed by the internet is mostly sufficient for us to view. It is difficult when comes to information gathering and analyse from various web applications. For instance, when monitoring and analysing housing market, users need to open few web applications and manually compare the detail information site-to-site. It takes amount of time for the investors around the world to consider the residential properties that are difficult to access, because the experience of analysing housing market could not be applied in a different region

(Singla & Bendigari, 2019) According to the research, there are differences about the factors that affect house renting prices in a different region. (Capoche-Angerer, 2016)

Conclusion of the research shows that common factors such as number of bedrooms, number of bathrooms, pets allowed, square footage, available number of parking slots, washer/dryer, location and lawn are well-known of the 8 factors that may affect the rental prices. To find out the factor that affect the residential properties renting value of a region. One of the solutions is to collect list of residential information, perform compare and contrast analysis to understand the underlying dependency of attributes that affect the housing value.

In this fast pace era, ease of accessing to the internet makes all of us became the one who always observes information in a lightning speed. For instance, social media deliver the hottest news without needing to turn on the TV to watch the news (Martin, 2018). As adapting the concept of "quick observation", a dashboard will more likely be introduced for users to help them easily access the core information (Bakusevych, 2018). Ideally, user would monitor the information at a glance. Where the detailed information can be clearly monitored on a single interface and compare through visualised map to deliver efficient information that the users care the most. Hence, complex, and inconsistent information that are published from web applications can be summarised into a dashboard.

Preferably, a dashboard is expected to solve the problem of inefficient information delivered to the users. However, a well-established dashboard for housing market might have significant contribution to the society (Klipfolio.com, 2019). As the dashboard would help property investors and existing landlords to understand what is happening for the housing market within a certain region and to prevent renters to rent at higher price but return in low value.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This section would give a short review about the strengths, weaknesses as well as provide some proper recommendation for existing dashboards and web applications that is related to housing market. By reviewing the existing works, developer would get to know what the minimal and core functionality are to be included for the dashboard design, so that the dashboard to be delivered can be implemented to solve real-life problem.

(Few, S, 2004) Some dashboards are being miss-diagnosed or misinterpreted as a dashboard, they fail to communicate and having an inadequate design for the regarding data. During this section, the existing platforms which are meant to convey information to their user, would be evaluated based on reliable sources and least amount of author opinion. For instance, the evaluation will be referring to related journals, books or websites. By reviewing the systems would allow the developer to see what the demands in the real world are, what the existing systems are lacking and its availability to make an improvement. This section would also allow the developer to understand the logical and physical design of the dashboard to monitor housing market. From where the developer would find novelty, identify the user needs, defining objectives and understanding the concept.

Besides, to understand the important factors for residential properties. This section will provide some examples and advices given by other researchers. As it may help the developer to conduct usable and meaningful dashboard rather than focus only on the design of the dashboard. Moreover, it may also help the audience to understand how the author perform data understanding during this section.

2.2 Data Source

2.2.1 Mudah.my

Mudah.my is an online marketplace where provides connection to buyers and sellers in Malaysia. From the introduction of the official website, “All Malaysian can sell and buy something on their website” (Mudah.my, n.d.). Mudah.my allow users to upload

residential property for renting and selling purpose, where user can make offer by contact the seller. They provide search bar to search for the requested item name and filter bar to find the items as users' preferences. The advertisement for Mudah.my quote that "Search and purchase nearly anything in Malaysia". Tons of data uploaded by users can easily be found by using Mudah.my. In terms of houses for rent in Kampar, Mudah.my have over 60 results listed from Jun until August in the year of 2019. The searched result contains attributes such as image, area/apartment name, price, unit type, house or room size (square feet), number of bedrooms, and number of bathrooms. From where to assume these are the attributes that peoples from Mudah.my think are important to convey.

Strengths

Mudah.my provide clean interface for user to view. Search bar and filter features are useful for items searching. Besides, Mudah.my counts and show the number of requested items found for users. This information would be useful when comparing Mudah.my with other online marketplaces to consider which platform has the most results found.

Weaknesses

Most of the "house for rent" results found in Kampar does not provide the exact location or address. The information described by sellers may not be reliable or detailed. For instance, seller may write "5 minutes away from McDonald". Therefore, instead of information telling by words, location plotting on a map will deliver more appropriate and descriptive information that tenants might want to know.

2.3 Existing Dashboard

2.3.1 MarketMinder

MarketMinder is a web-based service provided by AirDNA.co to deliver dashboards for all Airbnb rental worldwide. According to Few, S. (2004) 'A dashboard is visualising the core information that is needed to achieve multiple objectives.' MarketMinder delivered 13 KPIs as its most important information. In short, this report would focus on evaluating the first four dashboards which displayed on top of the

website. They are average daily rate, occupancy rate, revenue and market grade. Except market grade, each of the dashboards highlighted the highest and lowest value, the exact rate of the finding would be indicated apparently on the left of the graph with large font size. From where users can easily monitor the information at a glance. However, users could view the detailed findings from the inner dashboard by clicking the hyperlink below the dashboard. All the attributes except "Market Grade", consist of simplified linear visualisation. "Market Grade" highlighted the overall evaluation of the Airbnb market performance, concluded by a grading five aspects of the region, including "Rental Demand", "Revenue Growth", "Seasonality" and "Regulation". Every dashboard description was described clearly in a textbox on the top-right corner of the dashboard, it can be viewed by move the cursor on top of information button to avoid chaotic dashboard layout.

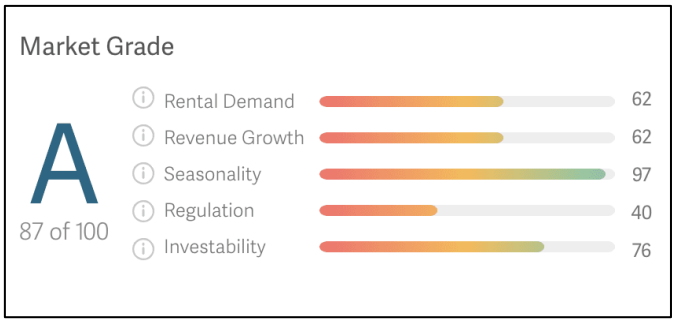


Figure 2-1 Market grade dashboard provided by AirDNA (Anon., 2018)

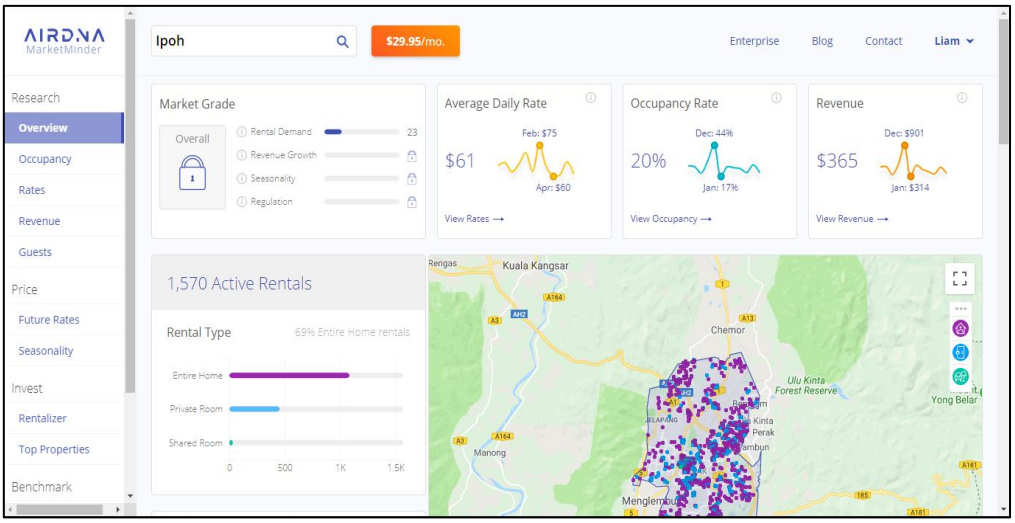


Figure 2-2 Airbnb house renting performance dashboard by requested region developed by AirDNA (OptimizeMyBnb.com, 2018)

Strengths

MarketMinder provide a clean interface with the white background and neat dashboard layout arrangement. The proposed dashboard allows important information to be monitored at a glance. By minimising the linear relationships, the dashboard able to attain simplicity. Besides, mapping out the information using scatter plot allowed users to gain information and insight within the requested region efficiently.

Weaknesses

The overall arrangement of dashboards can be improved to monitor information at a glance. Few, S. (2006) carried out that information being spotted within eye span, allow to make comparisons and spot relationships professionally. The problem of scrolling down would not only make us wonder the information lies below, the short-term memory could not help us to hold chunks of information. Therefore, comparing from dashboard-to-dashboard would not be efficient and the difficulty of insights finding arise. Moreover, the provided analysis services are not fully free to access. Therefore, normal user must purchase to fully access the analysis result.

2.3.2 Lancashire County Council Housing Dashboard

Lancashire county council housing dashboard (Anon., 2019) provides median houses price for year ending September 2018 in Lancashire with a total of 14 regions. The total of four dashboards composite the full image of Lancashire county council housing, including “Select District”, “Median house price for year ending Sep 2018 by local authority”, “Median house price for year ending September 2018 by ward” and “Deprivation and median house price relationship”. The provided “Select District” is a drop-down wizard which it would present a list of regions for user to choose for understanding the detailed visualisation of a city or multiple cities rather than general view on the county. “Median house price for year ending Sep 2018 by local authority” visualised as horizontal bar graph in descending order, it is list of static values where it could not be affected by changing the “Select District” field. Besides, the “Median house price for year ending September 2018 by ward” demonstrates the average house price for cities in the region, and under/over average price among England, represented

as Red (above England average) and Blue (below England average). A scatter diagram is used to picture the "Deprivation and median house price relationship", from where the IMD rate can be viewed for every city of the selected district.

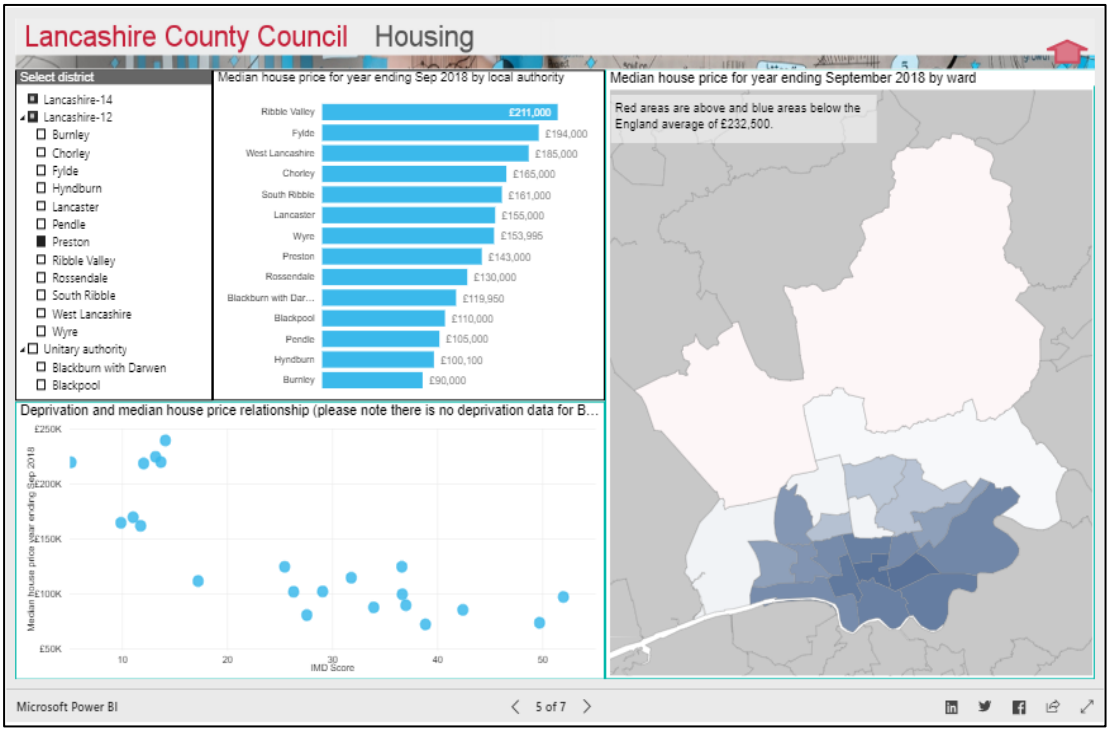


Figure 2-3 Lancashire county council median house prices dashboard for year ending September 2018 (Anon., 2019)

Strengths

Data collecting and cleaning phases are often challenging for a big dataset. Plot from the "Deprivation and median house price relationship" and "Median house price for year ending September 2018 by ward" can easily interact with each other. An interactive dashboard would show the detailed information of the requested attribute, for user to view and easy to understand the information delivered by the dashboard as well as gain insights.

2.4 Factors affecting rentals of residential apartments

Singla & Bendigari (2019) conducted a study show that every region would have different factors that affect the house rental price. The following table refers to what author claim as factors that affecting the rental price in Pune, India. The table provides

the positive relation with rental price and negative relation with rental price; The former says the price would increase as the factor rose; While the latter represent the price would decrease as the factor rose. This table may be used for comparing to the findings on the core factors that affect Kampar houses rental value. Since Kampar and Pune is in two different country where Kampar located far away from Malaysia capital and Pune is Maharashtra's culture capital in India. Both regions' population have big gap where Pune has estimated population of 3.99 million and (Citypopulation.de, n.d.) Kampar has estimated population of 0.1 million in the year of 2010.

Positive relation with rental price	**Negative relation with rental price**
Size in square footage	Safety factor: Distance from fire and rescue department
Number of rooms	Convenient related factor: Distance from employment zones, school/college and public transport terminal
Distance between central business district and hospitals	

Table 2-1 Factors that affect rental prices in Pune, India

2.5 Regression Models

One of the future works that is mentioned in FYP 1, is to build a machine learning model that can predict the residential property's renting price in Kampar. To investigate which of the model is suitable for this project. (Kojouharov, 2017) It is suggested by expert to attempt for these two models, Elastic Net - Lasso and Radial Basis Function. Both models are regression models that may be works on this project. To understand the models, this literature will be the studies about the two models. The models will be used if parameters RMSE and R-square value of the models is just right to the testing data. (Martin, n.d.) RMSE, also known as square root of the variance residuals, it is an absolute measure of fit, differences from R-squared, the relative measure of fit. R-

squared also known as the coefficient of determination, it is a measurement on how fitted is the data to the regression line, its value ranged is from 0%-100% (0% indicates the model shows none of the response data are surrounded its mean; 100% indicates all the response data are surrounded its mean).

2.5.1 Elastic Net – Lasso

Lasso

The lasso technique resolves this regularisation problem. Given Lambda value, lasso as the nonnegative value, solves the problem. The figure 2-4 illustrate as Lambda value increases, Beta, as the number of nonzero value decreases. The contrast of elastic net algorithm is Lasso problem includes L1 norm of Beta.

$$\min_{\beta_0,\beta}\left(\frac{1}{2N}\sum_{i=1}^{N}\left(y_i-\beta_0-x_i^T\beta\right)^2+\lambda\sum_{j=1}^{P}|\beta_j|\right).$$

Figure 2-4 Definition of Lasso (Anon., n.d.)

Elastic net

The elastic net is one of the techniques resolve the regularisation problem. It tunes the Alpha value to be approximately 0 and 1, and a nonnegative value of Lambda solves the problem. When Alpha value equal to 1, the Elastic net will be same as lasso. As Alpha value approaching 0, elastic net will be approaching ridge regression. For values that are than Alpha, penalty term P will be introduces between L1 normalisation of Beta and squared of L norm of Beta.

$$\min_{\beta_0,\beta}\left(\frac{1}{2N}\sum_{i=1}^{N}\left(y_i-\beta_0-x_i^T\beta\right)^2+\lambda P_\alpha(\beta)\right),$$

$$P_\alpha(\beta)=\frac{(1-\alpha)}{2}\|\beta\|_2^2+\alpha\|\beta\|_1=\sum_{j=1}^{p}\left(\frac{(1-\alpha)}{2}\beta_j^2+\alpha|\beta_j|\right).$$

Figure 2-5 Definition of Elastic Net (Anon., n.d.)

(Anon., n.d.) Lasso is a regularisation technique. It is used to reduce the number of attributes in a regression model, identify for the important attributes. Theoretically, elastic net is used when there are several highly correlated variables are existed in the dataset, however, due to the unknown understanding to the dataset, it can be used to identify whether it is the model to predict the renting price. When it is applying to unknown data, lasso estimator will produce smaller mean squared error than least-squares estimator. One of the differences between ridge regression and lasso is when the penalty grew, there will be more zero coefficient being set by lasso. A hybrid ridge of regression and lasso regularisation, elastic net, is similar to lasso, it can reduce model's size by generate zero values coefficients.

2.5.2 Radial Basis Function

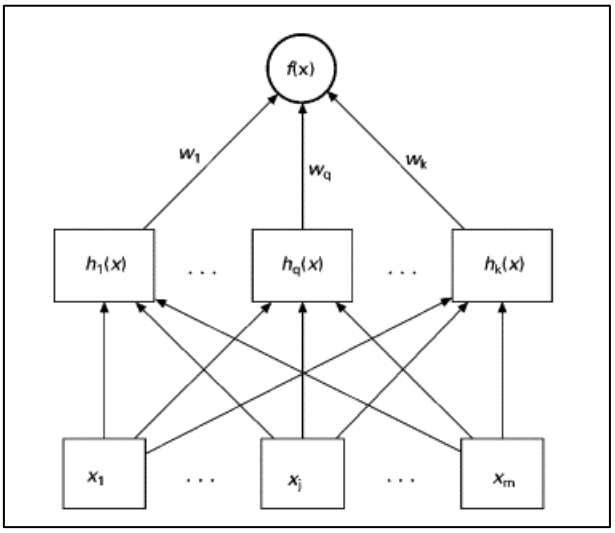


Figure 2-6 Architecture of Radial Basis Function Network (Faris, et al., 2017)

(Faris, et al., 2017) Radial basis function used for approximate multivariable functions by linear combinations of scenarios using a univariable function. It will only be used when there are known at a finite number of points in approximate functions. It can be applied in neural networks and learning theory. The proposed researched on the renting price prediction using the dataset may be suitable to this model, because this machine learning model can lies in its applicability in almost any dimension, it is then applicable without any knowledge on how many dimension on the dataset will be generated. Radial basis function network is a three-layer feedforward neural network. It contains, one or multiple hidden layers, single output layer, and a pass-through input layer. The process of transforming function into hidden layer also known as radial basis function.

CHAPTER 3: SYSTEM DESIGN

3.1 Overview of the Dashboard

3.1.1 Use Case Diagram

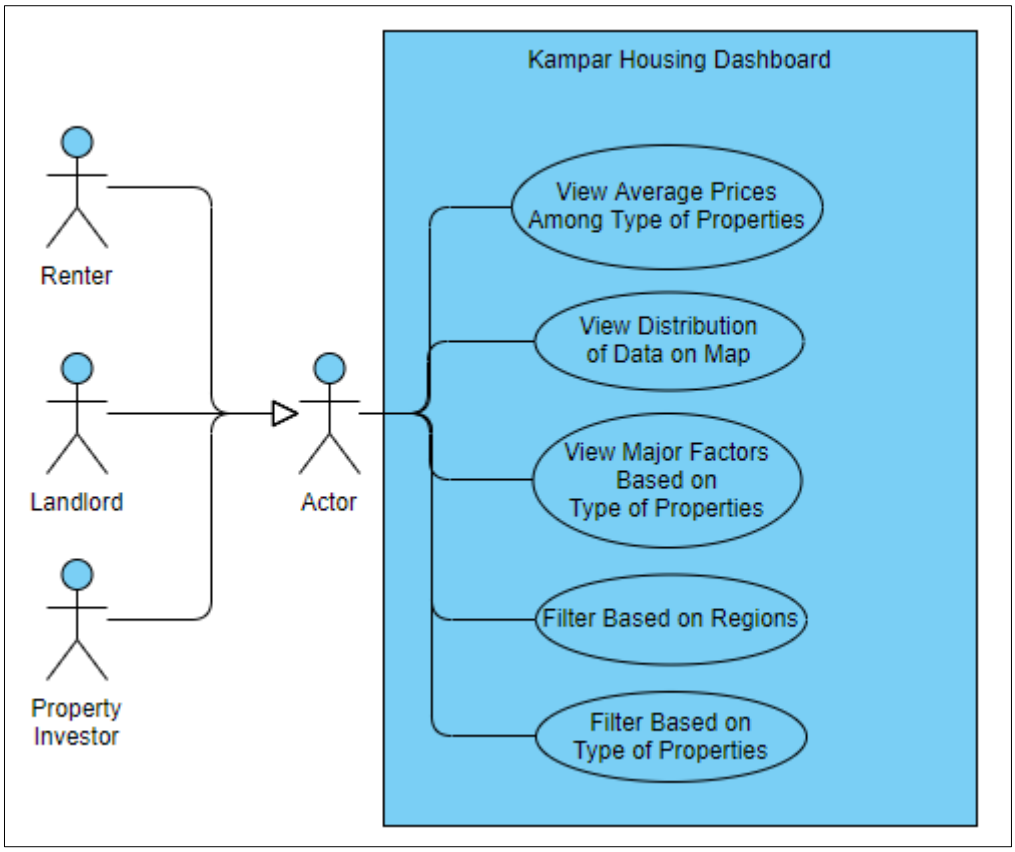


Figure 3-1 Use case diagram for the proposed dashboard

The illustrated use case diagram Figure3-1 shows that, users of the dashboard will be renters, landlords and property investors. The users can view and filter the dashboard based on the type of property to compare the renting price among the same group of properties within the 17 regions. From where the user can analyse the information conveyed by the dashboard and having a better understanding about the factor that affecting the renting price of the property.

3.1.2 Activity Diagram

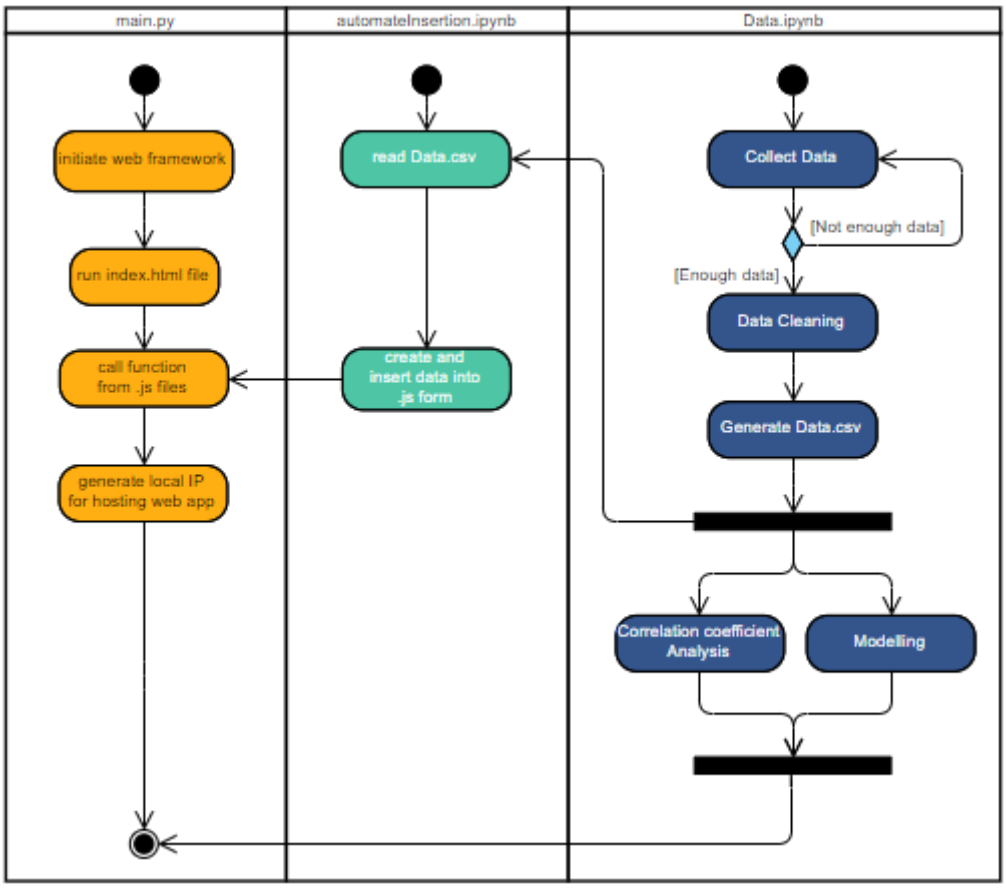


Figure 3-2 Activity diagram for the proposed dashboard

Figure 3-2 illustrate the full procedure to conduct the project. First, web scrapping will be performed to collect data from abstracting web applications. The collect data phase, also involving conduct survey. If the data does not meet the requirement, it is recommended to continue to collect data. When the number of data meet the requirement, which means data cleaning is now ready to be processed. In data cleaning phase, removed noise and fill null values with selected method will be done. Finally, the Data.ipynb will generate Data.csv for correlation coefficient analysis (to find the core factor that affecting residential properties renting price), modelling (to come out with a prediction model for renting price) and being read by automateInsertion.ipynb (to automate the process of inserting data into .js form). When the backend process is ready, run main.py, in this program, it will initiate a web framework, and call the function from .js files, including the output of automateInsertion.ipynb. Javascript files

which are not generated from automateInsertion.ipynb, are manual coded, they are mainly codes on calling Mapbox API, Chart.js API and adjusting aesthetics setting on them. The main.py will generate an IP address for hosting the web application.

3.2 Timeline

The figure 3-3 illustrates the planning timeline for each phase in FYP 2.

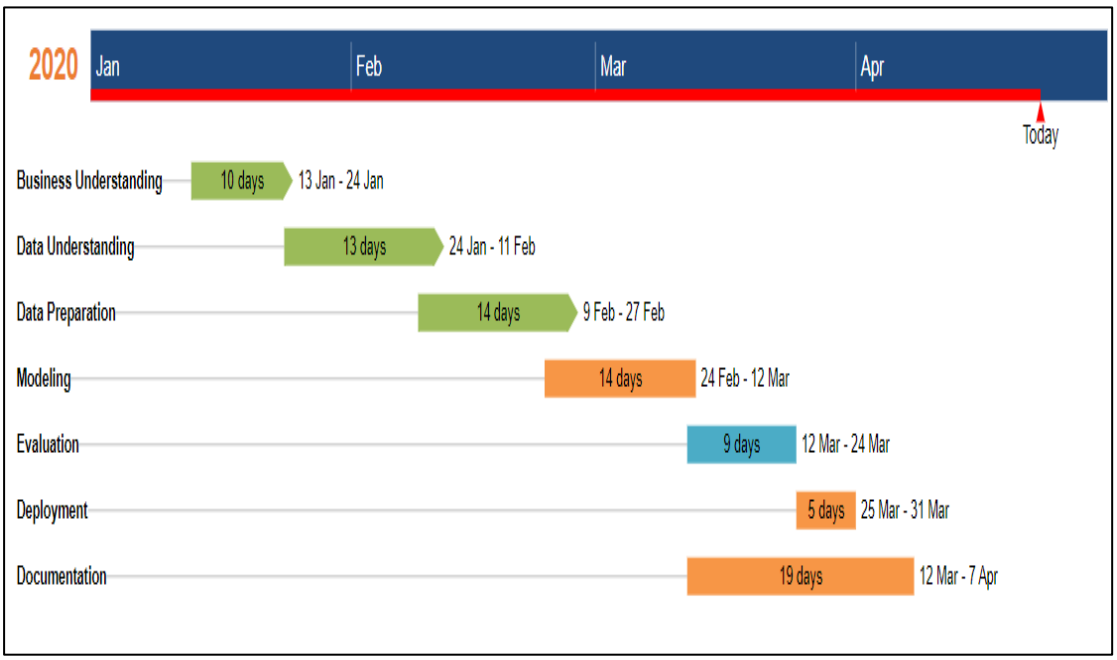


Figure 3-3 Timeline for FYP 2

CHAPTER 4: PROPOSED METHOD/APPROACH

4.1 CRISP-DM Approach

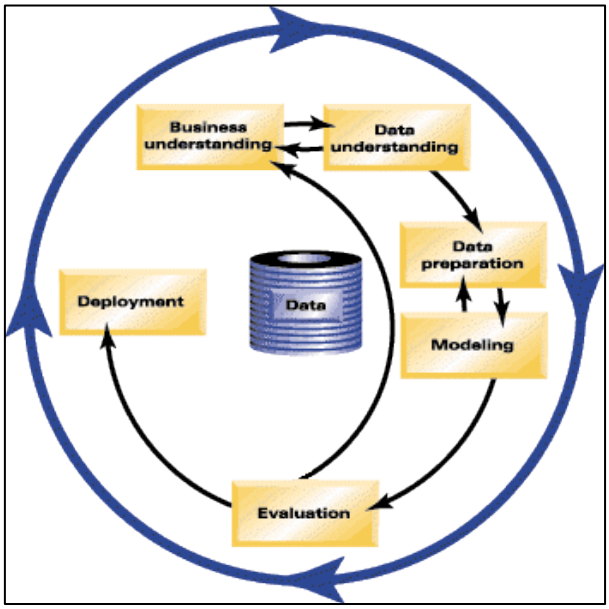


Figure 4-1 CRISP-DM (IBM, n.d.)

The chosen methodology to develop the dashboard is CRISP-DM methodology. (Wirth and Hipp, 2000) suggest a standard process model for data mining. CRISP-DM provides an overview for the data mining life cycle. Where the arrows indicating the most regular and core dependencies between phases (IBM, n.d.). This model has an advantage on its flexibility where the developer can 'modify the process phases by moving back and forward between phases if necessary' (IBM, n.d.). For instance, to develop a dashboard to monitor housing market would focus more on data collecting and visualisation. The reason of choosing this methodology is because the housing market dashboard will be needed a batch of data provided by various existing web applications. The data collecting process, however, would need the developer to perform data mining on the existing web applications and distribute survey to students from UTAR, Kampar campus. As a standard process of data mining, the most popular one also known as CRISP-DM.

4.1.1 Business Understanding

(SmartVisionEurope, n.d.) During business understanding phase, the developer needs to understand the purpose of developing the dashboard, define what are the existing

problems for analysing information from existing web applications. The developer needs to identify the problem statements, from where the objectives would be focusing on find the solution for the problems.

In this case, the solution will be focus on is to solve the problem of inefficient analysis process by switching to compare information in various websites, and vague information that web applications provide to users. Besides, it is recommended to understand the existing challenges faced by the developer such as the resources availability. For instance, this project would be focus on scraping any available data from web applications which can be implemented into the dashboard, and data mining tools or relevant software need to be determined in this phase. The data mining tool be used in this project is Beautiful Soup, which it can abstract data from html and xml mark-up languages. The relevant software be used is Anaconda, which is designed to help on simplifying package management.

4.1.2 Data Understanding

Data understanding mainly focus on helping the developer to recall the highlighted information, and as a guidance throughout the data mining process. This phase would involve in tools, methods, data source acquired for collecting data and encounter the problems throughout the data mining process. Preparing the dataset for the future development process.

For this project, data mining method will be web scraping and conducting survey to collect data from web applications and local university students. The labels for finding core factors that affecting renting prices and for renting prices prediction will be abstracted from web application data. (Singla & Bendigari, 2019) According to the research, every region having different factors that may affect residential properties renting price. With the investigations, this project has assumption on data which is uploaded to the web application from admins, are considered as the factors affecting properties renting prices, and whether the assumptions is true or false, the correlation coefficient and the prediction model will have the prove on it.

Data mining tools will be used to abstract data that are uploaded to the web application, and BeautifulSoup will be introduced to assist the developer to perform web scraping towards mark-up languages. After data is collected, the developer can explore the data to overview whether there is null value or get to know each attributes' datatype. Lastly, get the dataset ready to be clean in the next phase.

4.1.3 Data Preparation

(SmartVisionEurope, n.d.) This phase would be based on the data that is prepared in the previous phase and perform data cleaning. To utilise all the attributes from the dataset that may contribute to the project, some attributes need to be included or excluded, it is depends on the project scope and objectives of the project. Data preparation phase would also need the developer to tidy up the dataset. For instance, missing value could be ignored, adding assumption value or estimating value to fill in or to be discarded. In this project, Pandas would be introduced to conduct data cleaning process as the project is written in Python, Pandas also known as one of the powerful and popular tools to perform data cleaning.

4.1.4 Modelling

(SmartVisionEurope, n.d.) In this phase, developer may need to select a model which is best meet the objectives that are identified in the first section. There are two main categories involved in determining which model to be use for the problem, known as regression, and classification problem. During this phase, the dataset would be partitioned into two parts, testing set and training set. The training set will be the dataset to fit into the selected model, and the testing set will be the validation set to test the performance of the model before it moved forward for evaluation. It is recommended to compare different models and their performance. In this project, it is identified as a regression problem.

Based on the research, this project is suitable to apply regression model, for instance, Radial Basis Function, Ridge Regression, or Lasso Regression to predict the residential properties renting price in Kampar. Throughout the process of searching the most

suitable model, it can conclude that Radial Basis Function will be the most suitable model among the three, it produce the minimum mean square error, however, with very limiting data access, the model will not be able to have a sustainable outstanding performance. (Chernick, 2012) It is suggested to have minimum of 210 samples for each class, the ideal minimum samples for the prediction to renting price will be having an approximately 10710 samples in total (210 samples for every class including 17 regions and 3 type of residential properties), because of having large amount of observations, it is safe to say having more data will have more opportunities to reduce the mean square error.

4.1.5 Evaluation

During the evaluation phase, the selected model would be revised to identify whether the selected model is suitable to be used, and whether it meets the project objectives. If the selected model is not suitable for the case, the developer may need to determine the weaknesses or drawback of using it. To determine a proper model to be use, the developer need to review and recall the objectives and the whole process. Making sure that each process is met the project objectives. Then, the developer would have choices regarding to the next phase. Determining whether the status is ready for deployment or take another round of iteration until it met the expectation.

In this phase of study, the developer found that the size of data must be greatly increases. Although the current dataset seems to be enough for the prediction, however, due to variation of classes are existed in the datasets, it may need more data in each of the classes to increase the reliability and sustainability on the result. As research has been done in modelling phase, the evaluation result currently showing Radial Basis Function bringing the lowest mean square error.

4.1.6 Deployment

After the evaluation phase is done, the result may bring into deployment. The developer needs identify which way of presenting the result is the best and suitable for this project

and produce the final report as well as final presentation. In this project, dashboard would be used to visualise the result of data collecting. At the end, the developer may record down the experience gained from this project, any obstacles and pitfall that the developer successfully encountered.

The main obstacle on this project is the limitation of accessing data source. Although seeking for data from internet and distribution of survey seems to be a reliable solution to encounter the problem, however, the solution might have to be improved rather than seeking data from e-commerce websites, it is possible to seek data from social media since there are over millions of activities are happened in social media every day. This project is data dependent; therefore, it is recommended to seek for more data to deliver reliable message through dashboard.

4.2 Tools

Google Maps Geocoding API

After data is collected, it is found that residential properties that are listed in web applications did not have exact latitude and longitude. Therefore, Google Maps Geocoding was introduced to smoothen the process of finding the latitude and longitude based on the properties name listed in the web applications.

Mapbox API

The project is currently using default setting of the map provided by Mapbox. The previous solution on scatter visualisation, is using Plotly library to plot data on the selected area that are defined by latitude and longitude attributes, however, it is found that Plotly have limited customisation to the Mapbox aesthetics and access to its API to deliver a customised dashboard, as well as having limited sources and experiences been found on the internet. Therefore, the project is then moved from Plotly, to Javascript in delivering more aesthetic methods and provide visualisations flexibility.

Flask

The developed dashboard is considered to deployed as a web application. Therefore, Flask will help to initiate the web framework. It would provide appropriate libraries, technologies, and tools to improve the web application functionality. As an open source web framework, its popularity makes the development process easy and documentation of it can be easily be accessed from the internet.

Anaconda Navigator

Anaconda Navigator is a platform that controls and manages packages in this project. From where Spyder IDE, Jupyter Notebook, Python and other required packages versions can be monitored and modified through the platform.

Spyder IDE

(Anon., n.d.) Spyder is an integrated development environment that used, designed, and developed by group of scientists and data analysts. It allows cross file editing, debugging, profiling and analyse for a development tool with data evaluation. Besides, Spyder also offering built-in integration with most scientific packages, they are NumPy, Pandas, Matplotlib and more.

Jupyter Notebook

(Anon., n.d.) Jupyter Notebook is an open-source web application that allows this project to create and share the result of the research. It allows live code demo, visualisation, and narrative text to the audiences. Is a good presentation tool in this project.

4.3 User Requirement

The users for this project are renters, landlords and properties investors. It is set as the targets for the proposed dashboard to be achieved. In real situation, the project will be

request from client, vendors must discuss with the client about the user requirement and conclude it as a user story. The below are the user story conducted for this project according to existing dashboard and integrate with the objective of this project.

- As a renter, I want to have an overview on Kampar properties renting price so that I can rent the property with a reasonable price.
- As a renter, I want to have a better overview on analysing Kampar properties rather than analysing from multiple web applications.
- As a landlord, I want to have an overview on Kampar properties renting price so that I can improve my property value.
- As a landlord, I want to know what price range of the property within one area is.
- As a landlord, I want to know what is surrounding the property that I rent.
- As a property investor, I want to have an overview on Kampar properties renting price so that I can purchase a higher returns property in Kampar.
- As a property investor, I want to filter the type of property to compare which type of property having higher return on investment.
- As a property investor, I want to know what is surrounding the property that I rent.

4.4 Hardware and Software Specification

TYPE	SPECIFICATION
Hardware	Intel Core i5-4200 CPU, 2.8GHz
	64-bit Operating System
	Installed memory (RAM): 8GB
	Hard Disk Space: 480GB SSD
	Graphic Card: GEFORCE 820M

Table 4-1 Hardware specification

TYPE	SPECIFICATION
Software	Python 3.6.10
	Spyder 4.0.1
	Conda 4.8.3
	Jupyter notebook 6.0.3

Table 4-2 Software specification

CHAPTER 5: VERIFICATION PLAN

After the dashboard has been developed, in this project, it is suggested to perform testing on the usability aspect, which is non-functional testing. In this section, it is found that usability testing is sufficient to cover the dashboard. The sub-testing item on usability testing will consider visibility and usability matter, this is because the main function of this dashboard, is to deliver at a glance and interactable module.

5.1 Usability Testing

1. Visibility of Dashboard Page			
Purpose: The proposed dashboard should keep user informed, through appropriate feedback when performing an action.			
Step No.	**Review Checklist**	**Yes / No / Not Applicable**	**Comments**
1.1	Are all the font design and stylist dealing consistent across the entire page?	Yes	All fonts that are on the page are kept consistent.
1.2	Are all the fonts on the page displayed clearly without over wrapping the screen?	Yes	All the fonts is big enough, easily to be seen and read without any difficulty.
1.3	Does the system notify user when interaction is performed on the page?	Yes	Mouse cursor style will be changed when cursor hovering a clickable object.

Table 5-1 Visibility Testing on Dashboard Page

2. Usability of Dashboard Page			
Purpose: The proposed dashboard should allow users to interact and use the features available.			
Step No.	**Review Checklist**	**Yes / No / Not Applicable**	**Comments**

2.1	Can the user click and view each type of residential properties information upon clicking cardboard placed on top of the page?	Yes	Users can click every cardboard and view the information about region in both map and radar chart.
2.2	Can the user click listing region and view the region information in detail?	Yes	Users can click every listing on the side bar and view the region information in the block that is placed at top right corner of the map.
2.3	Can the user click and view the distribution of the plotting?	Yes	Users can click the merged plots to view the exact distribution of the plots.
2.4	Can the users click on the same cardboard twice times to show and hide the merged/unmerged plots?	Yes	Users can click on the same cardboard twice times to show and hide the merged/unmerged plots.
2.5	Can the users click on the same listing region twice times to focus and back to original view?	Yes	Users can click on the same listing region twice times to focus and back to original view.
2.6	Can the users click on the legend on radar chart to hide and show plotting on radar chart?	Yes	Users can click on the legend on radar chart to hide and show plotting on radar chart.

Table 5-2 Usability Testing on Dashboard Page

CHAPTER 6: CONCLUSION

6.1 Novelties, Contributions and Problems Encountered

6.1.1 Open Source Dataset

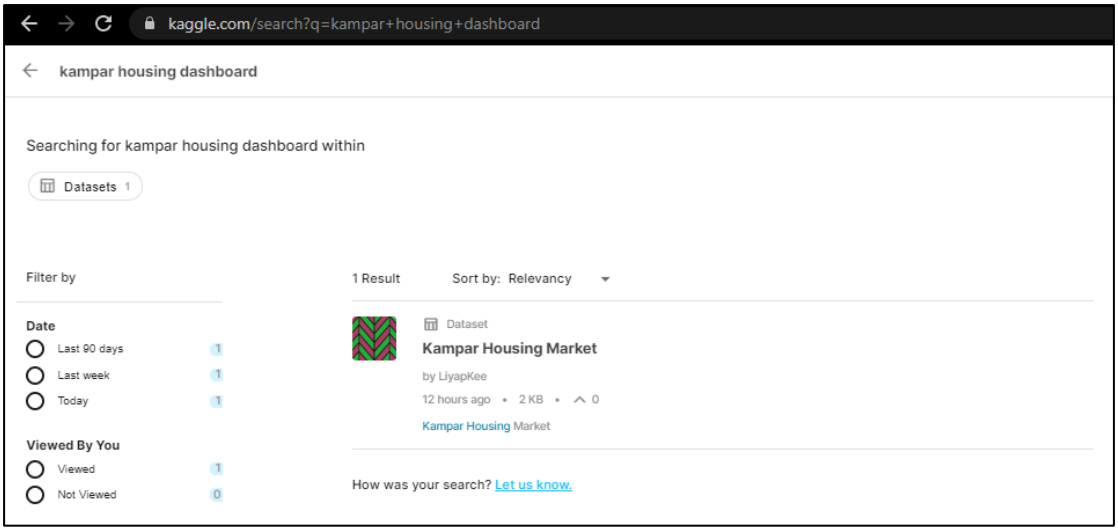


Figure 6-1 Uploaded dataset on Kaggle

Throughout the project development, scatter data that are from various web applications have been combined using web scraping tools and survey conducted. The main contributor of the web applications to the dataset is Mudah.my, which contribute around 30 sample to the dataset. Besides, due to the insufficient of data which have been collected from web applications, the developer decided to conduct survey and distribute among students from UTAR, Kampar campus, this survey helps to collected 92 set of sample, uses total of 138 sample to present on the dashboard after removing unused and bias data. It resulted in enough data to be visualised in the dashboard. (Yap, 2020) Finally, the collected data is uploaded to Kaggle as a contribution to the future works and having a chance to improve the dataset in different degree as open source dataset.

6.1.2 Factor that Affecting Kampar Residential Properties Renting Price

Figure 6-2 illustrate the correlation coefficient toward factors and house renting price. The items lie on y-axis, are the abstracted factors. And the x-axis indicates the correlation coefficient. The maximum correlation coefficient towards house renting price is the number of bedrooms.

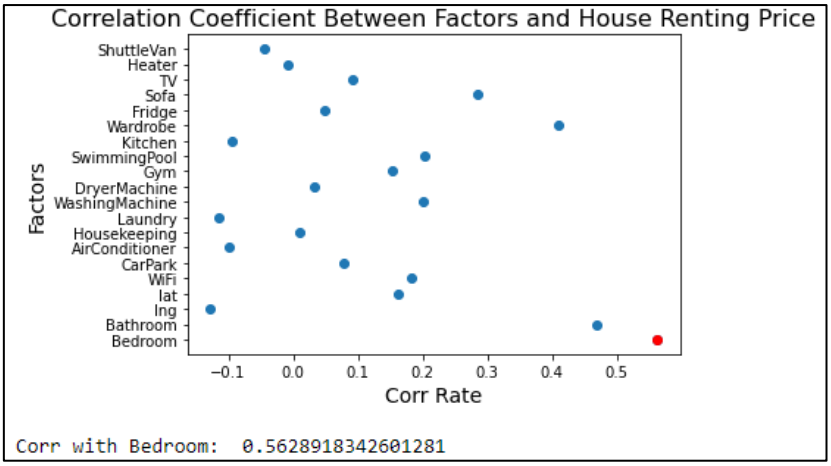


Figure 6-2 Correlation coefficient between factors and house renting price

Figure 6-3 illustrate the correlation coefficient toward factors and apartment renting price. The items lie on y-axis, are the abstracted factors. And the x-axis indicates the correlation coefficient. The maximum correlation coefficient towards apartment renting price is the number of bedrooms.

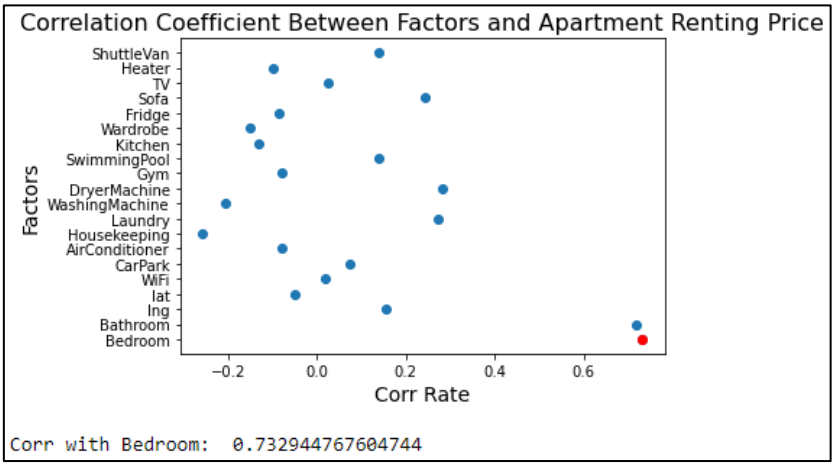


Figure 6-3 Correlation coefficient between factors and apartment renting price

Figure 6-4 illustrate the correlation coefficient toward factors and room renting price. The items lie on y-axis, are the abstracted factors. And the x-axis indicates the correlation coefficient. The maximum correlation coefficient towards room renting price is whether there is air conditioner in the room.

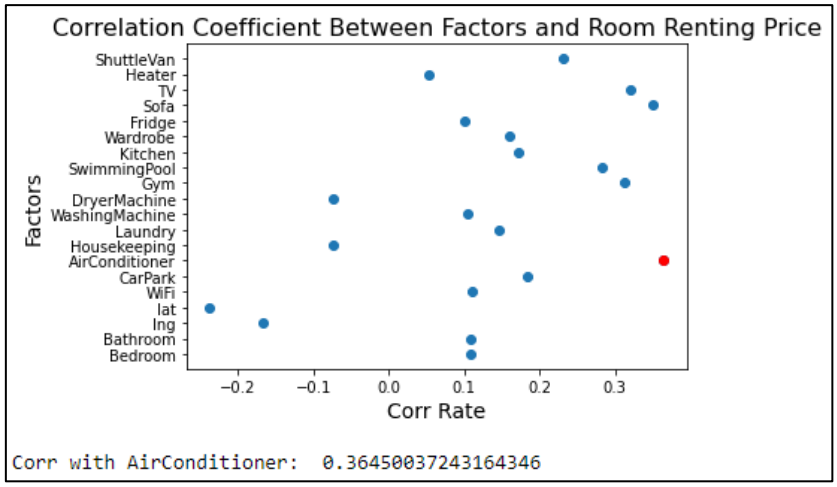


Figure 6-4 Correlation coefficient between factors and room renting price

The method of finding the core factor affecting Kampar residential properties renting value is using Spearman coefficient correlation among attributes from the collected dataset, it resulted in approximately 0.56 and 0.73 as the maximum coefficients correlation compare with number of bedrooms to house renting price and number of bedroom to apartment renting price accordingly; Nevertheless, the research shows the maximum value of coefficients correlation for room renting price is air conditioner, having approximately 0.36 correlation. Although the coefficient correlation in this research shows the relationship between renting prices and the targeted factor is weak, it is mainly because of variation of factors are monitored in this project, as for now, the maximum result is chosen as the most effective factor to the renting prices (correlation coefficient that is close to 1 or -1, leads to high relationship). Number of bedrooms is the core factor that affecting house and apartment renting price and renting price of room may be affected by whether there is air conditioner in that room.

6.1.3 Prediction on Kampar Residential Renting Price

The figure 6-5 illustrated the performance of using Elastic net lasso to predict the renting price. It shows the result of 0.1 for R-squared value, and RMSE for 552.25. Its performance can be considered as underfitting, which is, the model is not correctly predict most of the data.

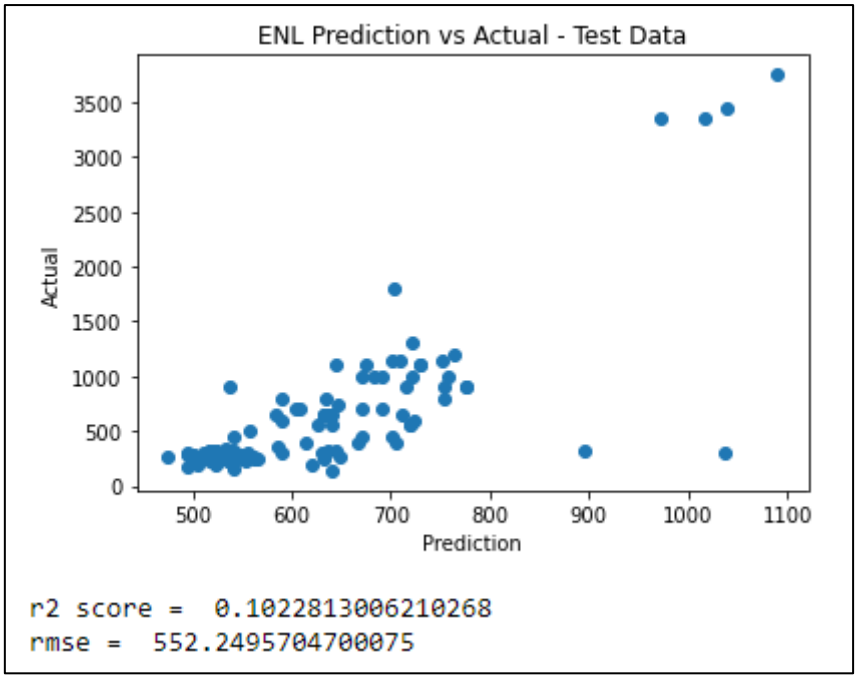


Figure 6-5 Elastic Net Lasso, prediction model on renting prices

The figure 6-6 illustrated the performance of using Radial basis function to predict the renting price. It shows the result of 0.84 for R-squared value, and RMSE for 40.49. Its performance can be considered low bias and low variance, which is the model, can predict most of the data approximate to the actual value.

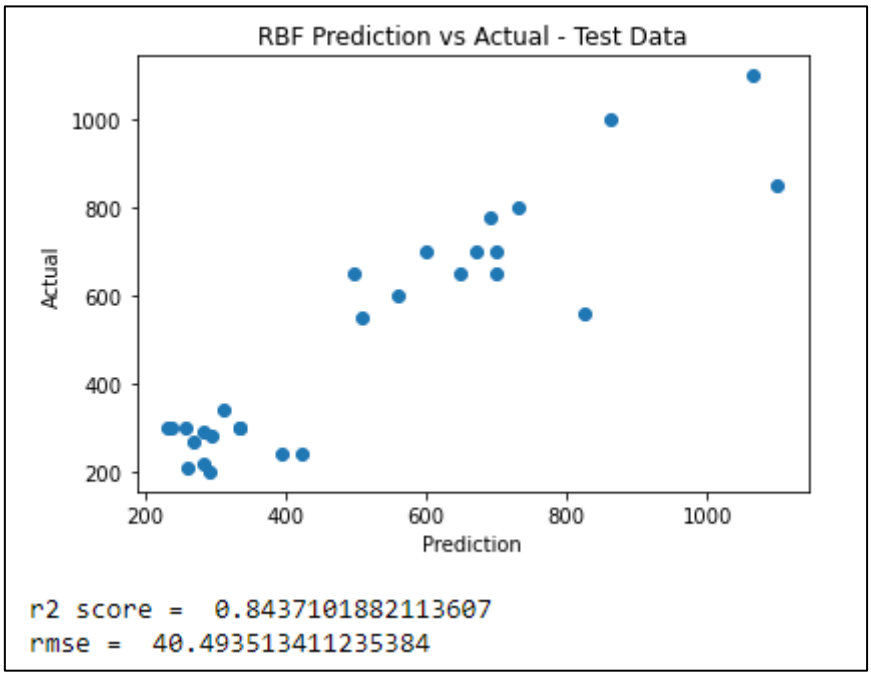


Figure 6-6 Radial Basis Function, prediction model on renting prices

Based on the collected data, this project has come out with a machine learning model to predict the residential properties renting price, the in-use model is radial basis function model. There are limitations while using this prediction model, it is only limited to predict the 17 regions in Kampar and the prediction performance is not good as expected. Due to the lack of data, the developer is not considering deploying the model to be used in the dashboard, until the model can be trained and adapt to sustainable high accuracy. Before that, it will be into the iteration and continue over again from collecting data.

6.1.4 Purposed Dashboard Overview

The figure 6-7 illustrate current dashboard overview.

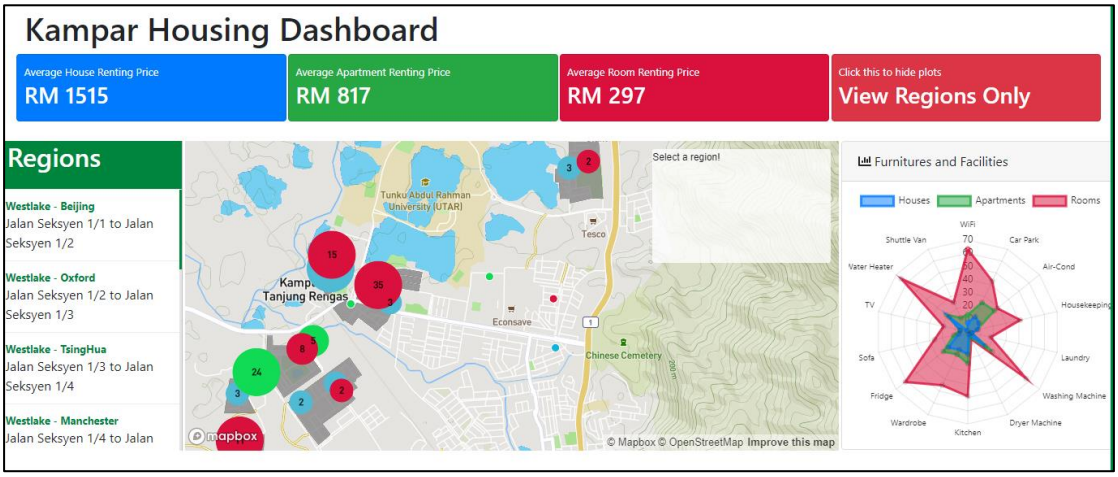


Figure 6-7 Kampar housing dashboard in FYP 2

The current dashboard user interface is clean and fit in page that allows user to gain information at a glance. There are total of 17 regions and 14 furniture and facilities are visualised through the dashboard. The features developed in this dashboard includes, show data distributions by type of residential properties, compare average renting price within regions, show particular region average renting price and exact location, and compare total furniture and facilities by type of residential properties. In this section,

features of the dashboard will be demonstrated and explanation of each features will be explained.

The figure 6-8 illustrate the view of cluster data is clicked or map is zoomed towards exact location of plotted data.

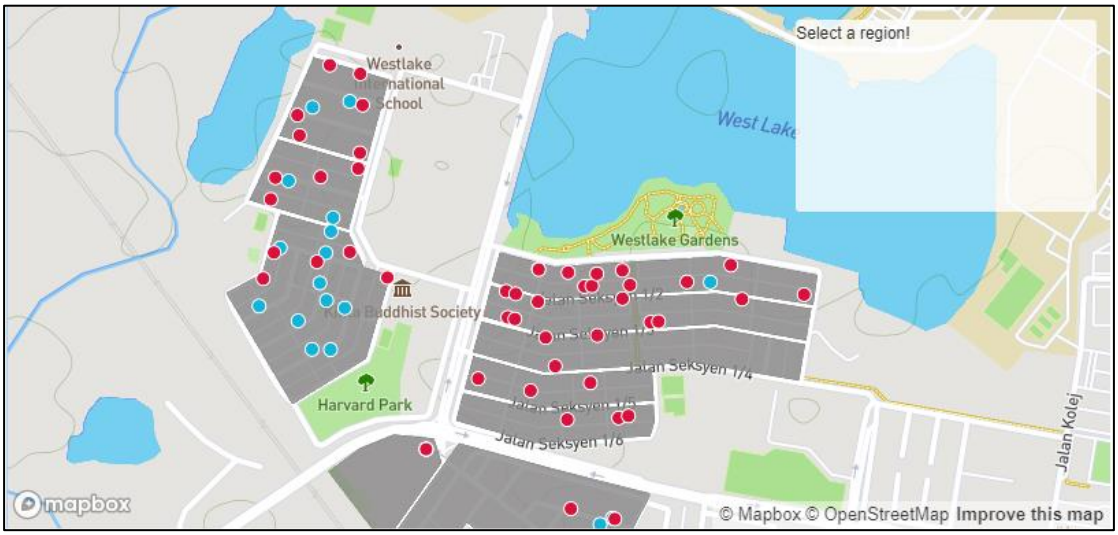


Figure 6-8 Exact location of each data

The first view on the dashboard is shown in figure 6-5, at this page, user can view all of the data plots on the map, it visualised the distribution of the data on the map, and the exact location of each plots can be view upon clicking the counted value of plot. Users can see where each data plotting exactly on the map as shown in figure 6-6. If users wish to back to the view as shows in figure 6-5, they are required to move cursor on the map and scroll to zoom in and out. User can also move around to view the non-merging plots.

The figure 6-9 illustrate after cardboard ‘Average House Renting Price’ is clicked.

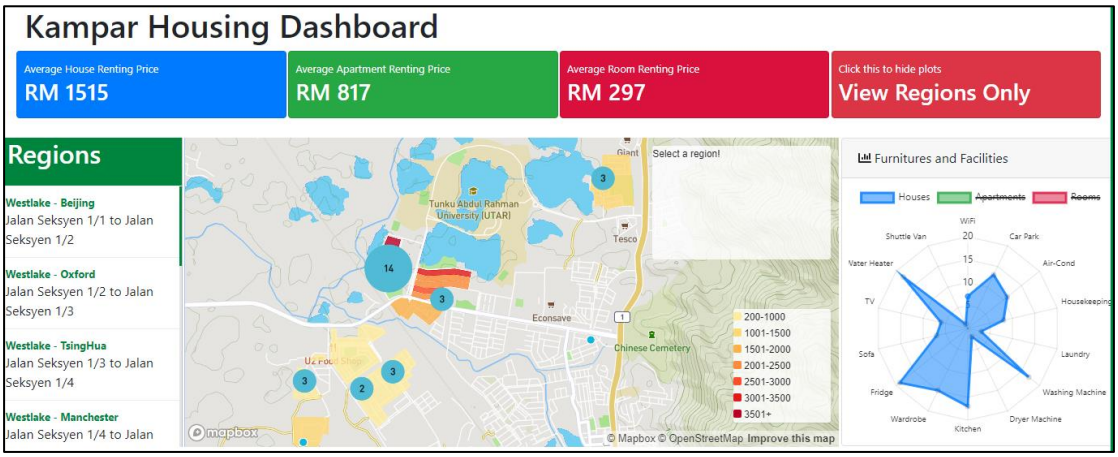


Figure 6-9 Average house renting price

The figure 6-10 illustrate after cardboard ‘View Regions Only’ is clicked.

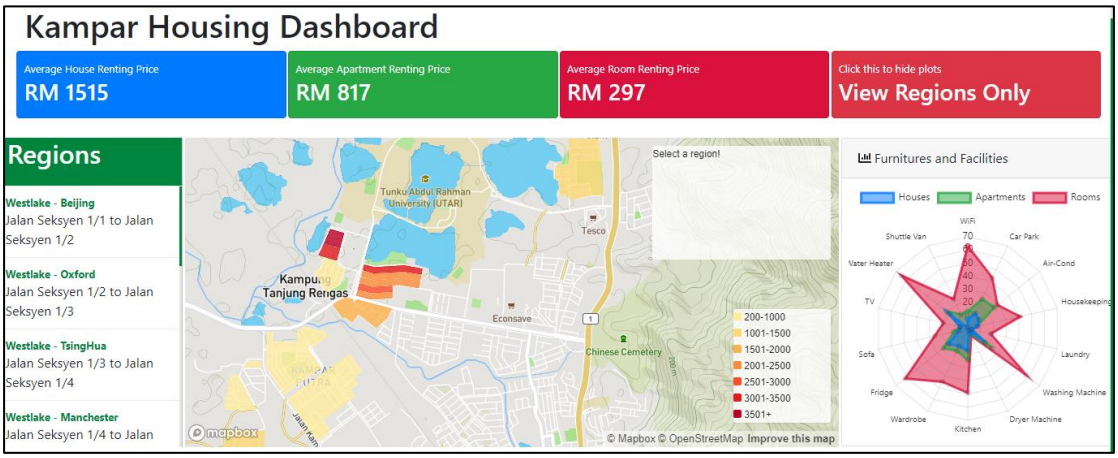


Figure 6-10 View regions

The cursor style will change when the user moves it on top of each cardboards that are coloured on top of the page. It will indicate each of the cardboards is clickable objects. The figure 6-7 demonstrates when users click on the ‘Average House Renting Price’ cardboard, notice that distribution of the other type of residential properties than houses will be hidden, and ‘Furnitures and Facilities’ radar chart will be showing data that are from houses type of residential properties only. Besides, each region will change colour to their respective average house renting price and the legend will be shown as a guidance to average prices comparison among regions. The same feature is applied for

cardboards ‘Average Apartment Renting Price’ and ‘Average Room Renting Price’. Next, cardboard ‘View Regions Only’, will hide the plots which is initially on the map. If users just wanted to compare among regions, click on ‘View Regions Only’ will hide all the plots that are initially on map and shows only regions information. The figure 6-8 demonstrate when ‘View Regions Only’ cardboard is clicked. Notice that after cardboard ‘View Regions Only’ is clicked, the radar chart will show all three type of residential properties information. This functionality is built because the developer wants to mention the comparison among three type of properties again for user to see the differences and understand the competitions as well as relations among these three type of properties, and it is not necessary to show all three type of properties distribution at once again in the dashboard, because the first of view is built for user to understand the overall data distribution only.

The figure 6-11 illustrate after listing regions ‘Westlake - Beijing’ is clicked.

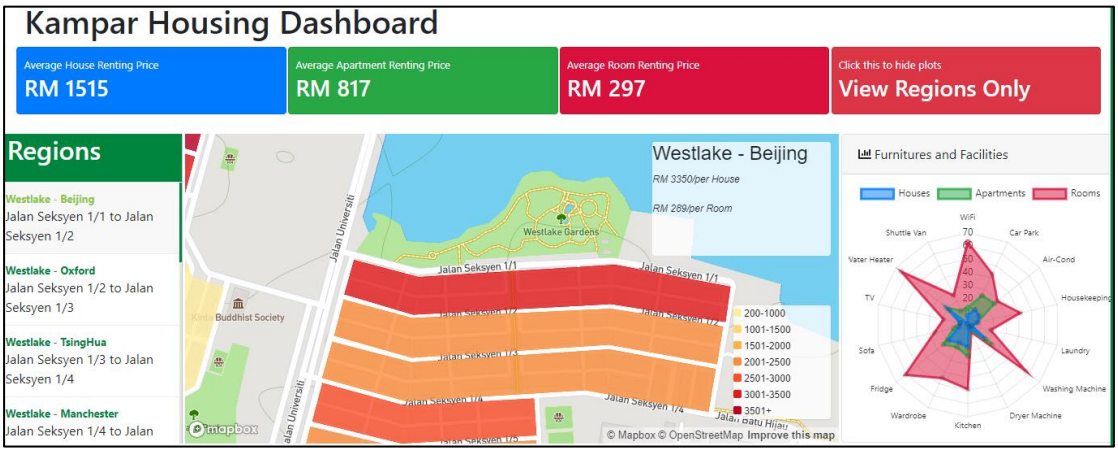


Figure 6-11 Select region

The left side listing the 17 regions information that are allowed user to click and focus the location on map. After select a region, the map will focus the exact location and shows average renting price for the selected region. The second click on the listing on the same region will return to the original view on the map. User can continually switch region to see each of the region’s information, the developed map will focus on the region according to user selection.

The figure 6-12 illustrate the ‘Furniture and Facilities’ radar chart.

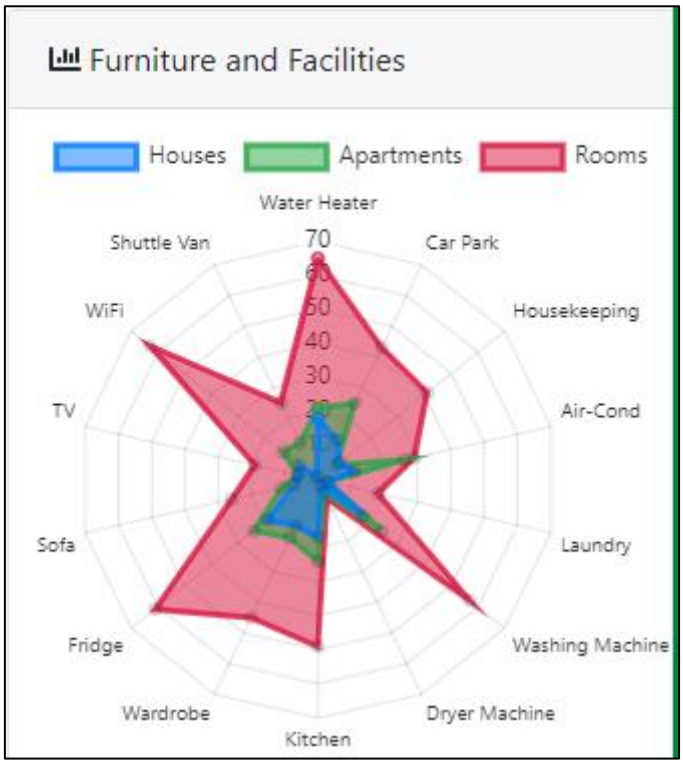


Figure 6-12 Furniture and Facilities Radar Chart

The developed radar chart as shown in figure 6-10, is presented 14 furniture and facilities labels. These labels are grouped by house (blue), apartment (green) and room (red) respectively. The visualised labels are to help users understand the competition on each group of the property types. Legend which is placed above the radar chart can be clicked and performed comparison with each and others. The radar chart tells the total number of each labels provided in this dataset for each type of residential property. Although the datasets having gymnasium and swimming pool, they are not considered as part of the labels in the radar chart. This consideration is because the above labels can easily tell user where to improve on, which is stated in the objective of the project.

6.2 Personal Insight and Future Work

From conducting the project, it provides the chance to search for the answer independently. Although the process is though and stressful, after overcome difficulty, it is found that nothing is impossible to be done. It brought full of joyfulness when there

are contributions really happening in this project. Apart from that, there are still possible to improve on the current result.

Data Size

The current data size is not suitable to perform any predictions, due to the variety of classes. It must be greatly increase or finding for any methods there can replace the data collecting in real life, for instance, over sampling can be perform for classification problems. If there are any methods regarding to something similar to over sampling on regression problems, it may implement into the current dataset.

Machine Learning and Correlation Coefficient

After the data size problem is solved, correlation coefficient result will be more reliable and helpful to the machine learning model's accuracy. From where produce more possible interaction with the dashboard, more contributions and insights will be share through the dashboard.

BIBLIOGRAPHY

Anon., 2018. *Analysis of MarketMinder.* [Online] Available at: https://optimizemybnb.com/airdna-review-marketminder/ [Accessed 15 August 2019].

Anon., 2019. *Housing dashboard | Lancashire.gov.uk.* [Online] Available at: https://www.lancashire.gov.uk/lancashire-insight/population-and-households/households-and-housing/housing-dashboard/ [Accessed 13 August 2019].

Anon., n.d. *CRISP-DM Help Overview | Ibm.com.* [Online] Available at: https://www.ibm.com/support/knowledgecenter/en/SS3RA7_15.0.0/com.ibm.spss.crispdm.help/crisp_overview.htm [Accessed 15 August 2019].

Anon., n.d. *Data Visualization - What it is and why it matters.* [Online] Available at: https://www.sas.com/en_my/insights/big-data/data-visualization.html [Accessed 15 August 2019].

Anon., n.d. *Docs.spyder-ide.org.* [Online] Available at: http://docs.spyder-ide.org/ [Accessed 21 April 2020].

Anon., n.d. *Guru99.* [Online] Available at: https://www.guru99.com/levels-of-testing.html [Accessed 21 April 2020].

Anon., n.d. *Jupyter.org.* [Online] Available at: https://jupyter.org/ [Accessed 21 April 2020].

Anon., n.d. *MathWorks.* [Online] Available at: https://www.mathworks.com/help/stats/lasso-and-elastic-net.html [Accessed 5 April 2020].

Bakusevych, T., 2018. *10 rules for better dashboard design.* [Online] Available at: https://uxplanet.org/10-rules-for-better-dashboard-design-ef68189d734c [Accessed 15 August 2019].

Capoche-Angerer, C., 2016. *8 Factors Affecting Rent Prices.* [Online] Available at: https://www.linkedin.com/pulse/8-factors-affecting-rent-prices-christina-capoche [Accessed 25 July 2019].

Chernick, M. R., 2012. *StackExchange.* [Online] Available at: https://stats.stackexchange.com/q/37835 [Accessed 21 April 2020].

Citypopulation.de, n.d. *Kampar (District, Malaysia) - Population Statistics, Charts, Map and Location.*. [Online]
Available at: https://www.citypopulation.de/en/malaysia/admin/perak/0810__kampar/
[Accessed 15 August 2019].

Drafor Amenyah, I. a. A. F. E., 2013. *Factors Determining Residential Rental Prices.* [Online]
Available at: http://www.aessweb.com/pdf-files/39-50.pdf
[Accessed 15 August 2019].

Europakonsument.at, 2017. *Comparison websites are often misleading | European Consumer Centre Austria.* [Online]
Available at: https://europakonsument.at/en/page/comparison-websites-are-often-misleading
[Accessed 15 August 2019].

Faris, H., Ibrahim, A. & Seyedali, M., 2017. Chapter 28 - Evolving Radial Basis Function Networks Using Moth-Flame Optimizer. In: *Handbook of Neural Computation.* s.l.:Acedemic Press, pp. 537-550.

Few, S., 2004. *Dashboard Confusion | Perceptualedge.* [Online]
Available at: https://www.perceptualedge.com/articles/ie/dashboard_confusion.pdf
[Accessed 15 August 2019].

Few, S., 2006. *Information dashboard design.* Sebastapol, CA: O'Reilly.

Hayes, B., 2019. *Programming Languages Most Used and Recommended by Data Scientists |.* [Online]
Available at: https://businessoverbroadway.com/2019/01/13/programming-languages-most-used-and-recommended-by-data-scientists/
[Accessed 15 August 2019].

Klipfolio.com, 2019. *Data Dashboards. Definition, Design Ideas plus 3 examples.* [Online]
Available at: https://www.klipfolio.com/resources/articles/what-is-data-dashboard
[Accessed 26 July 2019].

Kojouharov, S., 2017. *Cheat Sheets for AI, Neural Networks, Machine Learning, Deep Learning & Big Data.* [Online]
Available at: https://becominghuman.ai/cheat-sheets-for-ai-neural-networks-machine-learning-deep-learning-big-data-678c51b4b463
[Accessed 5 April 2020].

Martin, K. G., n.d. *Assessing the Fit of Regression Models - The Analysis Factor.* [Online]
Available at: https://www.theanalysisfactor.com/assessing-the-fit-of-regression-models/
[Accessed 5 April 2020].

Martin, N., 2018. *How Social Media Has Changed How We Consume News.* [Online]
Available at: https://www.forbes.com/sites/nicolemartin1/2018/11/30/how-social-media-has-changed-how-we-consume-news/#4a9f3d273c3c
[Accessed 15 August 2019].

Misirlakis, S., 2017. *The 7 Most In-Demand Programming Languages of 2018.* [Online]
Available at: https://www.codingdojo.com/blog/7-most-in-demand-programming-languages-of-2018
[Accessed 15 August 2019].

Mudah.my, n.d. *About Mudah - Mudah.my.* [Online]
Available at: https://www.mudah.my/about/index.htm?ca=9_s
[Accessed 15 August 2019].

OptimizeMyBnb.com, 2018. *AirDNA Review, Analysis of MarketMinder - OptimizeMyBnB.com.* [Online]
Available at: https://optimizemybnb.com/airdna-review-marketminder/
[Accessed 15 August 2019].

Ryan, H., 2018. *Australian House Prices Are Falling. Here's What That Means for Your Rent.* [Online]
Available at: https://www.buzzfeed.com/hannahryan/house-prices-rent-australia
[Accessed 15 August 2019].

Singla & Bendigari, 2019. *Factors Affecting Rentals of Residential Apartments in Pune, India: An Empirical Investigation | Emerald Insight.* [Online]
Available at: https://www.emerald.com/insight/content/doi/10.1108/IJHMA-12-2018-0097/full/html
[Accessed 15 August 2019].

SmartVisionEurope, n.d. *What is the CRISP-DM methodology?.* [Online]
Available at: https://www.sv-europe.com/crisp-dm-methodology/
[Accessed 15 August 2018].

Steer, D., 2018. *The Big Four Reasons Companies Struggle to Hire Data Talent.* [Online]
Available at: https://mode.com/blog/four-big-reasons-companies-struggle-to-hire-data-talent
[Accessed 15 August 2019].

Sullivan, J., 2018. *Data Cleaning with Python and Pandas: Detecting Missing Values.* [Online]
Available at: https://towardsdatascience.com/data-cleaning-with-python-and-pandas-detecting-missing-values-3e9c6ebcf78b
[Accessed 15 August 2019].

Wieinga, J. a. G. F., 2018. *Intro to Beautiful Soup.* [Online]
Available at: https://programminghistorian.org/en/lessons/intro-to-beautiful-soup#what-is-beautiful-soup
[Accessed 15 August 2019].

Wirth, R. & Hipp, J., 2000. *CRISP-DM: Towards a Standard Process Model for Data.* [Online]
Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.198.5133&rep=rep1&type=pdf
[Accessed 15 August 2019].

Yap, K. L., 2020. *Kaggle.* [Online]
Available at: https://www.kaggle.com/liyapkee/kampar-housing-market
[Accessed 21 April 2020].

APPENDIX A - REQUIRED DOCUMENTS

A.1 WEEKLY REPORTS

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: 3,3	**Study week no.:1**
Student Name & ID: Kee Li Yap 17ACB04196	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: A Dashboard to Monitor Housing Market	

1. WORK DONE

- Learn and Implement Beautiful Soup to perform web scraping.
- Use Jupyter Notebook to test the coding.
- Manage to get some data from websites.
- Scrapping are in progress.
- Contact some web application owner to request the access of API.

2. WORK TO BE DONE

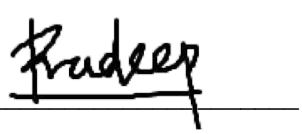
- Explore and confirm what tools to be use.
- Showing graphical visualisation on a local host.
- Get more data from multiple web applications.

3. PROBLEMS ENCOUNTERED

- Manage to perform web scrape using Beautiful Soup on one web application.

4. SELF EVALUATION OF THE PROGRESS

- Too nervous to learn causing slow learning progress.



Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: 3,3	**Study week no.:2**
Student Name & ID: Kee Li Yap 17ACB04196	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: A Dashboard to Monitor Housing Market	

1. WORK DONE

- Learn to use Dash for deployment.
- Moved the coding from Jupyter Notebook to Spyder.
- Explore tools to be use for visualising map.
- Contact some web application owner to request the access of API.

2. WORK TO BE DONE

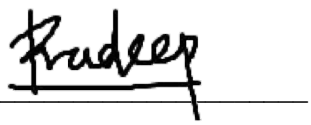
- Import the visualised map into web application.
- Get the latitude and longitude of every single data.
- Plot data on map using latitude and longitude as the x-axes and y-axes.

3. PROBLEMS ENCOUNTERED

- Manage to get more data from multiple web applications.

4. SELF EVALUATION OF THE PROGRESS

- Slow learning progress.
- Poor time management.



Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: 3,3	**Study week no.:3**
Student Name & ID: Kee Li Yap 17ACB04196	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: A Dashboard to Monitor Housing Market	

1. WORK DONE

- Moving from Dash implementation to Flask for deployment conveniences, open source features and documentations. As Dash is costly for more features.
- Use Mapbox API for importing graphic visualisation and Google Maps API to abstract latitude and longitude of the collected data.

2. WORK TO BE DONE

- Plotted data need to be more accurate.
- More data need to be collected to analyse the pattern.

3. PROBLEMS ENCOUNTERED

- Manage to immigrate smoothly from Dash to Flask before using too much of its functionality.

4. SELF EVALUATION OF THE PROGRESS

- Slow progress during immigration, code and documentation need to be rearrange.
- Poor understanding while reading online documentation.
- Poor information searching from search engine.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: 3,3	**Study week no.:4**
Student Name & ID: Kee Li Yap 17ACB04196	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: A Dashboard to Monitor Housing Market	

1. WORK DONE

- Setting the layout of the web application.
- Add colour bar to analyse the price range of the properties.

2. WORK TO BE DONE

- Find latitude and longitude of bus stops, McDonalds' and factors that may affect renting prices in Kampar.
- Plot the new collected data on the map to see the differences.

3. PROBLEMS ENCOUNTERED

- Resizing and repositioning legend.
- Manage to add colour bar to view properties' price range.

4. SELF EVALUATION OF THE PROGRESS

- Slow progress causing not much work done on the week.
- Must be more focus to get better result.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: 3,3	**Study week no.:5**
Student Name & ID: Kee Li Yap 17ACB04196	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: A Dashboard to Monitor Housing Market	

1. WORK DONE

- Collected latitude and longitude of the "hot spot" for visualisation purpose.
- Visualised new collected data to the map.

2. WORK TO BE DONE

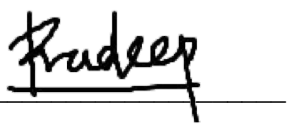
- Separate different type of property and using a legend to bind them.

3. PROBLEMS ENCOUNTERED

- Modify housing property name manually by reading property information posted by agents and using Google Maps to search for the nearest location of the properties.
- Property location is more accurate than before.

4. SELF EVALUATION OF THE PROGRESS

- Putting more effort than before, but still not enough.
- Push more on completing report.



Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: 3,3	**Study week no.:6**
Student Name & ID: Kee Li Yap 17ACB04196	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: A Dashboard to Monitor Housing Market	

1. WORK DONE

- Separate one trace into three traces to group properties' type.
- Add legend to separate three groups of properties and "hot spot"

2. WORK TO BE DONE

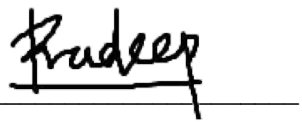
- Solve the contradiction of using the same colour / pattern to visualise the information that delivering by colour bar and legend.

3. PROBLEMS ENCOUNTERED

- Different traces allow the type of properties to be separated.

4. SELF EVALUATION OF THE PROGRESS

- Poor time management causing slow progress.
- Need to have more improve.



Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: 3,3	**Study week no.:7**
Student Name & ID: Kee Li Yap 17ACB04196	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: A Dashboard to Monitor Housing Market	

1. WORK DONE

- Modify and migrating to js form of document instead of python development.
- Append data to try out differences.

2. WORK TO BE DONE

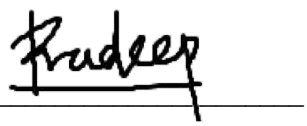
- Find a template for the website.
- Conduct survey to collect data.

3. PROBLEMS ENCOUNTERED

- Changes of codes is done, and migration is done smoothly.

4. SELF EVALUATION OF THE PROGRESS

- Poor time management causing slow progress.
- Need to have more improve.



Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: 3,3	**Study week no.:8**
Student Name & ID: Kee Li Yap 17ACB04196	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: A Dashboard to Monitor Housing Market	

1. WORK DONE

- Survey conducted.
- Template found and been used.

2. WORK TO BE DONE

- Data cleaning is ready for new dataset.
- Find the underlying relation between attributes and renting prices.

3. PROBLEMS ENCOUNTERED

- Template has been analysed and understood.

4. SELF EVALUATION OF THE PROGRESS

- Poor time management causing slow progress.
- Need to have more improve.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: 3,3	**Study week no.:9**
Student Name & ID: Kee Li Yap 17ACB04196	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: A Dashboard to Monitor Housing Market	

1. WORK DONE

- Data cleaning is done and been combined with old dataset.
- Reported correlation coefficient result into document.

2. WORK TO BE DONE

- Search for a model which is best fit for the problem.
- Understand and build a model to serve the problem.

3. PROBLEMS ENCOUNTERED

- Dataset is ready to be uploaded as open source dataset.

4. SELF EVALUATION OF THE PROGRESS

- Poor time management causing slow progress.
- Need to have more improve.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: 3,3	**Study week no.:10**
Student Name & ID: Kee Li Yap 17ACB04196	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: A Dashboard to Monitor Housing Market	

1. WORK DONE

- In progress with choosing a suitable model to be implement.

2. WORK TO BE DONE

- Models need to be selected.
- Include the literature review of models used.

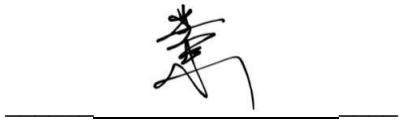
3. PROBLEMS ENCOUNTERED

- None

4. SELF EVALUATION OF THE PROGRESS

- Poor time management causing slow progress.
- Need to have more improve.

Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: 3,3	**Study week no.:11**
Student Name & ID: Kee Li Yap 17ACB04196	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: A Dashboard to Monitor Housing Market	

1. WORK DONE

- Models selected.
- Reported models architecture into literature reviews.

2. WORK TO BE DONE

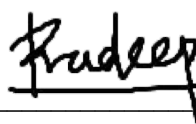
- Study for grid search and random search for fine tuning hyperparameters of the models.
- Write code to automate insertion of data into JavaScript file.
- Evaluate the performance of each models.

3. PROBLEMS ENCOUNTERED

- Dataset is uploaded to Kaggle.

4. SELF EVALUATION OF THE PROGRESS

- Poor time management causing slow progress.
- Need to have more improve.



Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: 3,3	**Study week no.:12**
Student Name & ID: Kee Li Yap 17ACB04196	
Supervisor: Dr Pradeep a/l Isawasan	
Project Title: A Dashboard to Monitor Housing Market	

1. WORK DONE

- Models are fine-tuned and been evaluated.
- Automated data insertion in JavaScript.
- Regions is drawn.

2. WORK TO BE DONE

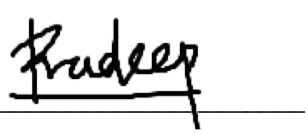
- All findings are ready to be well-documented.

3. PROBLEMS ENCOUNTERED

- Dashboard is successfully developed and deployed.

4. SELF EVALUATION OF THE PROGRESS

- Poor time management causing slow progress.
- Need to have more improve.



Supervisor's signature

Student's signature

A.2 POSTER

A DASHBOARD TO MONITOR HOUSING MARKET

Student & ID : Kee Li Yap 17ACB04196
Supervisor : Dr Pradeep a/l Isawasan

INTRODUCTION

This project is to propose a dashboard for data analysis on housing properties with the help of existing information publish on the internet.

PROBLEM STATEMENT

- The difficulty of analysing information from site-to-site.
- The vague information delivered by web applications.

PROJECT SCOPE

- Focus on collecting Kampar housing data from web applications.
- Develop a dashboard to visualise the factor effecting Kampar housing market.

PROJECT OBECTIVES

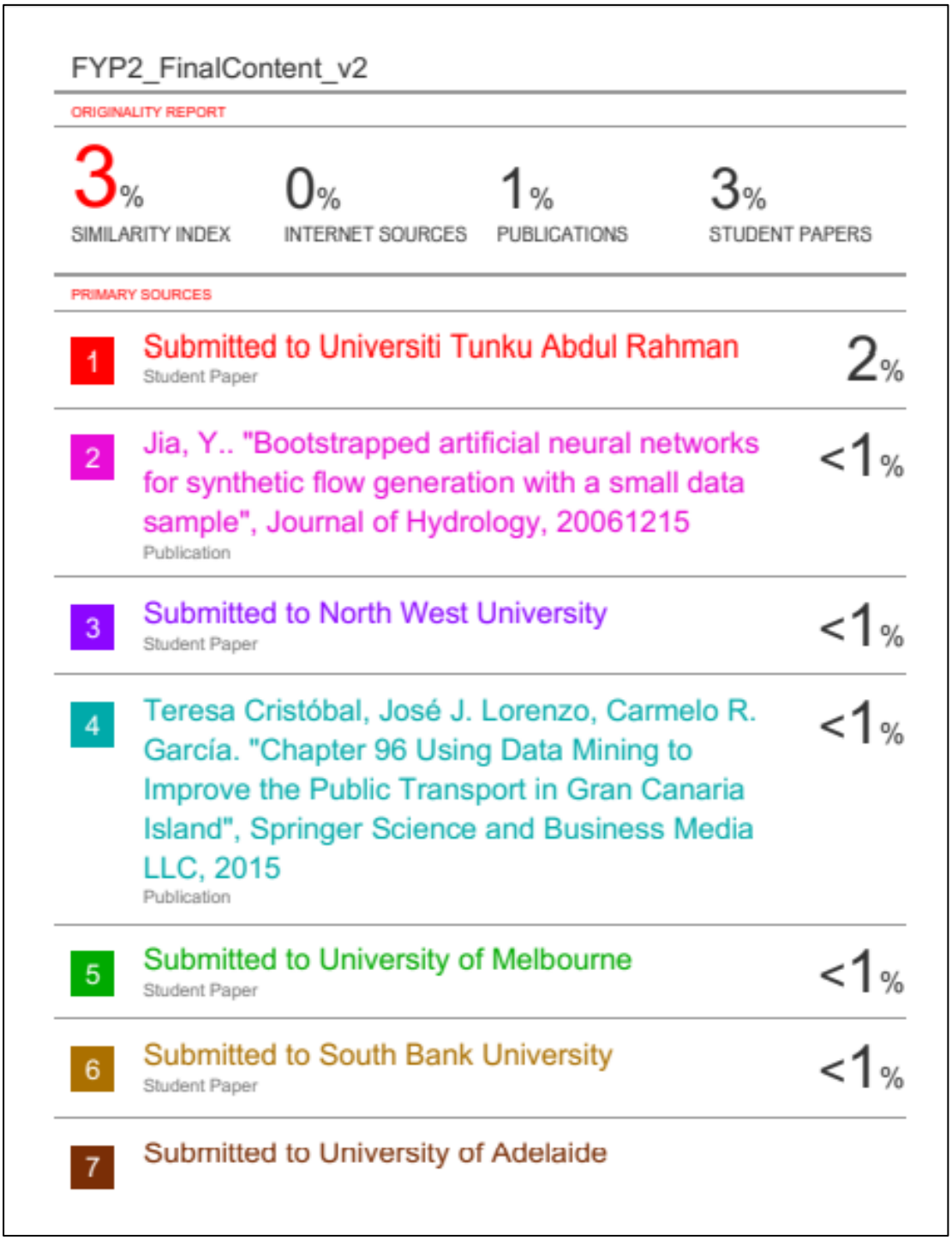
- To investigate the core factor that affecting rental prices of residential properties in Kampar.
- To develop a dashboard to monitor Kampar housing market.

CONCLUSION

The proposed dashboard can help renters, landlords to rent a property at a reasonable price and provide quick analysis for property investors.

UNIVERSITI TUNKU ABDUL RAHMAN

A.3 PLAGIARISM CHECK RESULT



FYP2_FinalContent_v2

ORIGINALITY REPORT

3%	0%	1%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universiti Tunku Abdul Rahman Student Paper	2%
2	Jia, Y.. "Bootstrapped artificial neural networks for synthetic flow generation with a small data sample", Journal of Hydrology, 20061215 Publication	<1%
3	Submitted to North West University Student Paper	<1%
4	Teresa Cristóbal, José J. Lorenzo, Carmelo R. García. "Chapter 96 Using Data Mining to Improve the Public Transport in Gran Canaria Island", Springer Science and Business Media LLC, 2015 Publication	<1%
5	Submitted to University of Melbourne Student Paper	<1%
6	Submitted to South Bank University Student Paper	<1%
7	Submitted to University of Adelaide	

	Student Paper	<1%
8	Submitted to University of Leeds Student Paper	<1%
9	Ruth A. Schmidt, Helen Wright. "Financial Aspects of Marketing", Springer Science and Business Media LLC, 1996 Publication	<1%
10	Submitted to Queen Mary and Westfield College Student Paper	<1%
11	Harish Kumar Singla, Priyanka Bendigiri. "Factors affecting rentals of residential apartments in Pune, India: an empirical investigation", International Journal of Housing Markets and Analysis, 2019 Publication	<1%
12	Submitted to Maastricht School of Management Student Paper	<1%

Exclude quotes	Off	Exclude matches	< 8 words
Exclude bibliography	Off		

Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 1	Effective Date: 24/04/2020	Page No.: 1of 1

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	**Kee Li Yap**
ID Number(s)	**17ACB04196**
Programme / Course	**CS**
Title of Final Year Project	**A Dashboard to Monitor Kampar Housing Market**

Similarity	**Supervisor's Comments** **(Compulsory if parameters of originality exceed the limits approved by UTAR)**
Overall similarity index: 3 % **Similarity by source** Internet Sources: **0 %** Publications: **1 %** Student Papers: **3 %**	
Number of individual sources listed of more than 3% similarity: **0**	
Parameters of originality required and limits approved by UTAR are as Follows: **(i) Overall similarity index is 20% and below, and** **(ii) Matching of individual sources listed must be less than 3% each, and** **(iii) Matching texts in continuous block must not exceed 8 words** ***Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.***	

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Signature of Supervisor

Name: PRADEEP ISAWASAN

Date: 24/04/2020

Signature of Co-Supervisor

Name: _____

Date: _____

UNIVERSITI TUNKU ABDUL RAHMAN

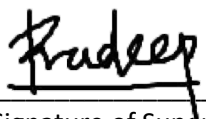
FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

A.4 - CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	17ACB04196
Student Name	Kee Li Yap
Supervisor Name	Dr Pradeep a/l Isawasan

TICK (√)	DOCUMENT ITEMS Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
/	Front Cover
/	Signed Report Status Declaration Form
/	Title Page
/	Signed form of the Declaration of Originality
/	Acknowledgement
/	Abstract
/	Table of Contents
/	List of Figures (if applicable)
/	List of Tables (if applicable)
/	List of Symbols (if applicable)
/	List of Abbreviations (if applicable)
/	Chapters / Content
/	Bibliography (or References)
/	All references in bibliography are cited in the thesis, especially in the chapter of literature review
/	Appendices (if applicable)
/	Poster
/	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)

*Include this form (checklist) in the thesis (Bind together as the last page)

I, the author, have checked and confirmed all the items listed in the table are included in my report.  (Signature of Student) Date: 24 APRIL 2020	Supervisor verification. Report with incorrect format can get 5 marks (1 grade) reduction.  (Signature of Supervisor) Date: 24/04/2020